Systematics and ecology of the brush-tailed mice, *Calomyscus*, in and around Pakistan based on the *Rbp3* gene

Research Thesis
Presented in partial fulfillment of the requirements for graduation
with research distinction in Biology in the undergraduate
colleges of The Ohio State University

by
Amanda Weller
The Ohio State University
December 2019

Project Advisor: Dr. Ryan W. Norris, Department of Evolution, Ecology, and Organismal Biology

Abstract

Family Calomyscidae is a group of muroid rodents consisting of the single genus *Calomyscus* with only 8 recognized species of brush-tailed mice. The genus is informally separated into three species groups, and my focus was on the southern *C. bailwardi* species group consisting of *C. bailwardi*, *C. hotsoni*, and *C. baluchi*. I analyzed how species present in this species group are related and investigated potential species suggested by previous studies. Interphotoreceptor retinoid-binding protein (*Rbp3*) sequences were obtained and a maximum parsimony network was constructed. The *Rbp3* network supports *C*. sp. Khy of northern Pakistan and *C*. sp. B of the Zagros Mountains in Iran to be new species, which had been suggested by *Cytb* and genomic analyses. The network less robustly supports separation of *C. hotsoni*, which shares a haplotype with one *C. baluchi* sample. Ecological Niche Modeling (ENM) indicates *C. baluchi* has a high affinity in its current distribution and has relatively good suitability throughout most regions where *Calomyscus* are generally known to inhabit. It also shows that *C. baluchi* has some suitability in the far north of Pakistan, but at lower elevations such as river valleys. The ENM model for *C*. sp. Khy indicates that this proposed species is highly specialized, with very little suitability outside where it was collected. The ENM model for *C. hotsoni* shows a general suitability throughout much of the projected distribution, although lower than *C. baluchi*.

Key Words: Calomyscidae, Balochistan, Hindu Kush, Ecological Niche Modeling

INTRODUCTION

South Central Asia sits at the interface of three biogeographic zones, the Palearctic, the Oriental, and to a lesser extent the Ethiopian (Roberts 1977). Pakistan and its surrounding regions exhibit extreme changes in climate and topography. Pakistan claims the fourth highest recorded temperature on the planet (Merlone et al. 2019) as well as regions of high snowfall, permanent snowfields, and below freezing temperatures (Roberts 1977). Elevation spans from sea level to high mountain ranges including K2, the second tallest mountain peak in the world. These extreme features of this region can influence species distributions.

Described by Thomas in 1905, *Calomyscus* are small rodents representing a very distinct evolutionary lineage having one of the earliest divergences from muroid rodents 25 million years ago (Schenk et al. 2013). They are the sole genus in the family Calomyscidae and are commonly referred to as brush-tailed mice due to their signature tufted tails (Norris et al. 2008). They are found in South Central Asia, from Syria to Pakistan west of the Indus River Valley (Kilpatrick 2017), and populations tend to be isolated due to uneven habitat. They are pale in color, have large ears, and prefer dry rocky mountainsides and steppe forests with the exception of one species (*C. hotsoni*), which prefers deserts with sparse vegetation (Roberts 1977; Kilpatrick 2017; Hamidi et al. 2019).

Currently only 8 species are recognized based on morphological characteristics and phylogenetic analysis of mitochondrial DNA (Musser and Carleton 2005; Kilpatrick 2017). Until Vorontsov et al. (1979), who described *C. urartensis*, all members of the genus were treated as subspecies of *C. bailwardi*, demonstrating the difficulty in classifying them based on morphology alone. *Calomyscus* can be informally separated into 2 or 3 species groups: the *incertae sedis C. tsolovi* from Syria with almost no information available; the *C. mystax* group

with 4 recognized species; and the *C. bailwardi* species group on which I focused my study (Kilpatrick 2017). The *C. bailwardi* group contains 3 of the 8 recognized species: *C. bailwardi* Thomas, 1905; *C. hotsoni* Thomas, 1920; and *C. baluchi* Thomas, 1920.

Calomyscus bailwardi is found in the Zagros mountains of Iran and lives in sub-alpine mountain forests or dry rocky mountainsides in areas of scarce vegetation (Eskandarzadeh et al. 2018). They are found between 400 and 3500 meters and are rarely encountered in low elevation valleys (Kilpatrick 2017). Their ears and hindfeet are typically similar length, and their tails are slightly longer than their head-bodies (Kilpatrick 2017).

Calomyscus hotsoni is found in the dry hot deserts of Iran and Pakistan between elevations of 67 to 1890 meters (Kilpatrick 2017; Hamidi et al. 2019). They are one of the smallest *Calomyscus* and their ears are typically smaller than their hindfeet (Kilpatrick 2017). Most *Calomyscus* have similar coat color, but *C. hotsoni* have black tipped hairs on their dorsal ends which helps in identification (Kilpatrick 2017).

Calomyscus baluchi is morphologically similar to *C. bailwardi*, although they have slightly smaller ears (Kilpatrick 2017). They are found in the rocky hillsides of Northern Pakistan and Eastern Afghanistan between the elevations of 400 to 3500 meters.

Studies done on this species group suggest the potential addition of 4 evolutionary significant units based on sequences of the cytochrome b (*Cytb)* gene and genomic studies (Norris et al. 2003, in prep; Akbarirad et al. 2016, 2018; Rawson 2019). Among these are *C*. sp. B, a group from the eastern Zagros mountains and *C*. sp. Khy, which were collected north of the Khurram River in in the Safed Koh mountain range and some of the Hindu Kush (Norris et al. 2003, in prep; Akbarirad et al. 2016; Rawson 2019).

I used the interphotoreceptor retinoid-binding protein (*Rbp3*) gene to analyze how two of the three recognized species in the *C. bailwardi* species group (*C. hotsoni* and *C. baluchi*) are related to each other and if there are biogeographical or ecological barriers that led to the diversification of two suggested new species (*C*. sp. B and *C*. sp. Khy; Akbarirad et al. 2015; Norris et al. in prep). I utilized Ecological Niche Modeling (ENM) to test the suitability of the ranges of these two potential species to investigate if there were any observable patterns to explain their apparent separation from other species. Unfortunately, no tissue was available for *C. bailwardi* or for the other new species suggested by mitochondrial data and these were not included in my *Rbp3* analyses.

MATERIALS AND METHODS

DNA extraction, sequencing, and alignment.— Tissue was provided by C. William Kilpatrick at the University of Vermont (UVM), originating from field expeditions led by Charles A. Woods and Kilpatrick from 1992-1998. Tissue was preserved in 95% ethanol and prior to the extraction process, it was soaked in distilled water to remove ethanol and improve yield (Kilpatrick 2002). DNA was extracted using the Gentra Puregene Mouse Tail Kit (QIAGEN; Germantown, MD). Locality information for samples, compiled from the Florida Museum of Natural History (FLMNH), are shown in Figure 1 and Appendix A.

The nuclear *Rbp3* gene was amplified using the primers IRBP119A2 and IRBP878F (Jansa and Voss 2000). PCR products were cleaned and sent to the TACGen sequencing facility (Richmond, CA). PCR products were directly Sanger sequenced by TACGen, which used the same primers previously used for PCR. Once raw sequences arrived back, I processed and aligned them by eye.

Phylogenetic analyses.— Maximum parsimony analyses were conducted using PAUP* 4.0a (Swofford 2002). The maximum parsimony tree (Figure 2) was constructed for the *Rbp3* gene. The tree was rooted using midpoint rooting because all available outgroup taxa are at least 25 million years divergent (Schenk et al. 2013). A midpoint rooted network was constructed to visualize results from parsimony analysis.

Ecological Niche Modeling analyses.— Species suitability was tested using ENM for 3 species, *C. baluchi*, *C. hotsoni*, and the suggested *C.* sp. Khy. *Calomyscus baluchi* and *C.* sp. Khy were tested to determine if they are ecologically distinct in addition to being distinct based on *Rbp3* results. Locality information for Iranian *C. bailwardi* samples were largely unavailable and are not distinguished from *C.* sp. B in databases. Iranian *C. hotsoni* could be included alongside Pakistani samples because of the work of Hamidi et al. (2019). Records of each species localities were collected from the FLMNH Mammal Collections Database (http://specifyportal.flmnh.ufl.edu/mammals/, accessed January 2019). Additional Iranian occurrences were obtained from previous studies (Akbarirad et al. 2016, Hamidi et al., 2019).

Ecological niche models were constructed in Wallace (v1.0.5), an R (3.5.1 (2018-07-02))-based graphical user interface (GUI) application (Kass et al. 2018a). The application employs Maxent (3.4.1) to project ENMs based on georeferenced data points (GPS coordinates) and estimates species suitability in areas based on resolution of 19 bioclimatic variables, imported from WorldClim (https://www.worldclim.org/, accessed Oct 2019). It is important to add that GPS coordinates were not taken by Kilpatrick and Woods at the time of capture in Pakistan. They were estimated based on location descriptions, but these may not be exact. Feng et al. (2019) produced a guideline for reporting ENM models that was utilized as much as possible to increase reproducibility. Occurrence models were measured in the current time

period. Habitat suitability for all species tested in this study were extended out over much of the distribution of the genus. The background data parameter was left at the default 10,000 samples. Parameters were adjusted to optimize complexity of models based on sample size and models could be evaluated with linear (L), quadratic (Q), and hinge (H) feature classes and regularization multipliers (Kass et al. 2018b)

For *C. baluchi*, 83 records were thinned to 31 localities after spatial thinning, removing all but one sample that were within 1 km proximity of another. WorldClim bioclimatic variables were loaded using map center 66.918, 30.114 at 2.5 arcmin resolution. Extent of points were buffered with point buffers by 0.8 degrees to resemble the extent of the distribution. The default 10,000 background points were randomly sampled. Occurrences were non-spatially partitioned with random *k*-fold with n=4 folds (Kass et al. 2018b). Feature classes chosen were L, LQ, H, and LQH with regularization multipliers of 1-3 with a step value of 0.5 generating 20 models. These models were evaluated based on the lowest omission rate (OR) and the lowest 10 percentile training presence threshold (avg.test.or10pct) was prioritized as the best model (Kass et al. 2018b). Three models were tied for the top, LQ_1, LQ_1.5, and LQ_2 with a value of 0.094. Due to this tie the highest average test area under the curve (avg.test.AUC) was chosen evaluated with the LQ_1 having the highest value at 0.869 (Kass et al. 2018b). Of the bioclimatic parameters, 10 had response curves (3,8,13,15,19,6,9,11,12; Figure 3b; Appendix C). The "cloglog" prediction output was chosen with no threshold and the map prediction was plotted (Merow et al. 2013; Kass et al. 2018b). The model was projected to a new extent and visualized in Figure 3.

For *C.* sp. Khy, 21 records were thinned to 6 localities after spatial thinning, removing all but one sample that were in 1 km proximity of another. WorldClim bioclimatic variables were

loaded using map center 71.197, 34.869 at 5 arcmin resolution. Extent of points was buffered with point buffers by 0.8 degrees to resemble the extent of the distribution. The default 10,000 background points was selected, but only 27%. (n=2674) were able to be randomly sampled. Occurrences were non-spatially partitioned with jackknife (k=n) option due to the low sample number. The only feature class suggested in the module guidance was L. Regularization multipliers of 1-3 were chosen with a step value of 0.5 generating 5 models. These models were evaluated based the OR and the lowest avg.test.or10pct value was prioritized as the best model (Kass et al. 2018b). All models were tied for the top with a value of 0.5. Due to this tie the highest avg.test.AUC value was chosen evaluated with the L_1 model having the highest value at 0.854 (Kass et al 2018b). Of the bioclimatic parameters, 6 had response curves (3,7,9,13,14,18; Figure 5b; Appendix C). The "raw" prediction output was chosen to reduce complexity with no threshold and the map prediction for relative abundance was plotted (Merow et al. 2013; Phillips et al. 2017). The model was projected to a new extent and visualized in Figure 4.

For *C. hotsoni*, 22 occurrence records (Hamidi et al. 2019) were loaded into Wallace. WorldClim bioclimatic variables were loaded using map center 62.04, 29.583 at 5 arcmin resolution. The default 10,000 background points were randomly sampled. Extent of points were buffered with point buffers by 1 degree to resemble the extent of the distribution. Occurrences were non-spatially partitioned with random *k*-fold with n=4 folds. Feature classes chosen were L, LQ, H, and LQH with regularization multipliers of 1-3 with a step value of 0.5 generating 20 models. These models were evaluated based the OR and the model with the lowest avg.test.or10pct value was chosen (Kass et al. 2018b). Two models were tied for the top, LQ_1.5, and LQ_2, with a value of 0.092 (Kass et al. 2018b). Due to this tie the model with the highest avg.test.AUC value was chosen, with the LQ_2 having the highest value at 0.712 (Kass

et al. 2018b). Of the bioclimatic parameters, 5 had response curves (5,6,9,12,15; Figure 5b; Appendix C). The "raw" prediction output was chosen with no threshold and the map prediction for relative abundance was plotted (Merow et al. 2013; Phillips et al. 2017). The model was projected to a new extent and visualized in Figure 5.

RESULTS

I recovered 18 *Rbp3* haplotypes from 30 individuals sequenced. The network is shown in Figure 1 and a midpoint rooted tree is shown in Figure 2. Two samples (1489 and 976) could not be analyzed as initially recovered due to ambiguities at informative sites. These may represent heterozygotes due to the nuclear nature of the *Rbp3* gene and its inheritance coming from both parents. This creates a potential to express two different base calls at the same site. I treated these samples as if they were each two separate sequences (A and B) at the two possible positions on the network.

The *C.* sp. Khy samples clustered together and there were no shared haplotypes with other species (Figure 1). The *C.* sp. B samples also clustered together and did not share any haplotypes with other species (Figure 1). These results support the hypotheses that these may be distinct species compared to other species tested.

The two *C. hotsoni* samples (1278 and 1416) share a haplotype with one *C. baluchi* sample (1260) which was collected in North Waziristan (Figure 1). This is the northernmost population of *C. baluchi* in my dataset and *C. hotsoni* are the southernmost species of *Calomyscus*, which may suggest that this is an old haplotype.

The projection for *C. baluchi* shows high affinity around Quetta and the nearby mountain ranges (Figure 3a). Suitability for *C. baluchi* across the geographic range of the *C. bailwardi*

species group is relatively good with some exceptions, including the far northern mountain ranges where *C*. sp. Khy is found as well as in low elevation deserts. Also, the current range maps of *C. baluchi* include the Hazarajat Plateau (Kilpatrick 2017), but this area seems to be unsuitable based on the projection for the species. There were 9 bioclimatic variables (3,8,13,15,19,6,9,11,12; Appendix C) that influenced the model (Figure 3b). Seasonal variables that affected the model were mean temperature of the wettest quarter, mean temperature of the driest quarter and mean temperature of the coldest quarter (O'Donnell and Ignizio 2012). Isothermality is useful for isolated populations, and it measures fluctuations in night to day temperature changes within a month relative to changes during the whole year (Nix 1986; O'Donnell and Ignizio 2012). Variables related to water and precipitation include annual precipitation (important when considering water availability), precipitation seasonality (coefficient of variation, which measures variability of precipitation), and precipitation of wettest month (where extreme precipitation affects distributions; O'Donnell and Ignizio 2012).

The projection for *C*. sp. Khy suggests a highly specialized distribution with very little suitability outside where it was collected (Figure 4a). Bioclimatic variables (3,7,9,13,14,18; Appendix C) that affected the model were isothermality, extreme temperature and precipitation measures including temperature annual range and precipitation of the wettest and driest months, and seasonal variables like the mean temperature of the driest quarter as well as the warmest quarter (Figure 4b; O'Donnell and Ignizio 2012).

The projection for *C. hotsoni* suggests they range across lower elevations (Figure 5a). It shows an overlap with the *C. baluchi* range, suggesting they should come in contact. The bioclimatic variables (5,6,9,12,15; Appendix C) indicate effects from anomalies like maximum temperature in the warmest month and the minimum temperature in the coldest month (Figure

5b; O'Donnell and Ignizio 2012). Other factors were precipitation seasonality (CV), which measures variability of precipitation, the seasonal mean temperature in the driest quarter, and annual precipitation which highlights importance of water availability (O'Donnell and Ignizio 2012).

DISCUSSION

Current understanding of the evolutionary history in *Calomyscus* is based entirely on mitochondrial data, which is inherited from the mother. Akbarirad et al. (2016, 2018) used the cytochrome oxidase subunit I gene (*CoI*) in addition to the *Ctyb* gene in their phylogenic analysis, yet both are of mitochondrial origin. Tracking only the matrilineal heritage can be problematic because males may exhibit a different life history and hybridization may cause a matrilineal lineage to persist. A review article conducted by Toews and Brelsford (2012) found 126 instances where mito-nuclear discordance was observed in recent studies. It is important to test for mitochondrial introgression with nuclear loci to include a gene that is not affected by this phenomenon and, unlike mitochondrial genes, nuclear genes are inherited from both parents. The *Rbp3* gene is of nuclear origin, hence this study.

Also, new developments in genomic assessment generate large amounts of data for relatively little cost and tissue material. A ddRADseq approach was utilized by Rawson (2019) due to its ability to provide good results for species that have do not have reference genomes (Peterson et al. 2012). By comparing genomic results to this single locus study, indicated species delimitations are strengthened (Rawson 2019).

For the suggested *C.* sp. B, the *Rbp3* results corroborate the distinction from other *C bailwardi* group species (*C. hotsoni* and *C. baluchi*) indicated by *Cytb* and genomic studies

(Akbarirad et al. 2016; Rawson 2019). Because I lacked the samples, I was not able to recover any *Rbp3* data from true *C. bailwardi* to compare. Future studies should include samples from this taxon, and I expect that the results will agree with the distinctions suggested.

For the suggested *C*. sp. Khy., the *Rbp3* network shows three haplotypes clustering together distinctly from other haplotypes (Figure 1). Due to the *Rbp3* results agreeing with the distinctiveness of *C*. sp. Khy and the Khurram River representing a geographic isolating barrier, ENM was used to investigate whether there were ecological requirements associated with the regions that would relate to this separation.

The Safed Koh mountain range lies just north of the Khurram river and is where some of the samples of *C*. sp. Khy were collected by Kilpatrick and Woods. North of the Khurram River and the Safed Koh mountain range lies the eastern Hindu Kush mountain range, representing the northernmost point of the *C. bailwardi* species group. The Hindu Kush lies on the western edge of the greater Himalayan mountain complex, which also includes the Karakorum and Great Himalaya of northern Pakistan. The elevation in this region drops drastically from tall mountains to deep valleys. The Köppen-Geiger climate classification in the Safed Koh range indicates a semi-hot arid climate, and in the Hindu Kush range indicates tundra climate (Peel et al. 2007; Beck et al. 2018). The potential for allopatric isolation in these high mountain ranges may have grown during the Last Glacial Maximum (LGM). Precipitation from monsoons contributed to expanding glaciers, leading to allopatric isolation of these high-altitude populations (Kamp and Owen 2011).

Calomyscus baluchi exhibits immense *Rbp3* haplotype diversity, and even has a shared haplotype with *C. hotsoni* (Figure 1). The Köppen-Geiger climate classification in *C. baluchi* habitat ranges from hot desert, cold desert, semi-hot arid, and cold arid climates, highlighting its

ability to adapt to different habitats (Peel et al. 2007; Beck et al. 2018). The semi-hot arid and hot desert climate in the Suleiman range in central Pakistan shows very high suitability for *C. baluchi* (Figure 3a).

The *Rbp3* network shows a shared haplotype between *C. baluchi* and *C. hotsoni* (Figure 1), which may be due to admixture between the two species. The suitability of *C. hotsoni* in its realized home range is comparable to areas that *C. baluchi* inhabit. The lack of *C. hotsoni* in the Suleiman range suggests that *C. baluchi* may be a better competitor for this habitat. The ENM projects their distribution patchily across their range and the prediction levels are very low (highest level 0.0007), suggesting that the parameters tested with ENM might not be the best indicators to test the range of this species.

Calomyscus hotsoni is unique in the arid area in which it lives. Köppen-Geiger climate classification shows similarities between hot dry deserts and the range of *C. hotsoni*, yet the hot dry desert classification extends into *C. baluchi* and *C. bailwardi* habitat (Peel et al. 2007, Beck et al. 2018). Climatic oscillations during the LGM caused deserts to expand (Kamp and Owen 2011), potentially enabling *C. hotsoni* to have a broader distribution during that time and making them more competitive for habitat with *C baluchi*.

When soil conditions are suitable *C. hotsoni* have been observed to dig burrows, which is unique in comparison to other *Calomyscus* who live and nest in rock piles (Roberts 1977; Hamidi et al. 2017; Kilpatrick 2017). Even when soil conditions are poor, *C. hotsoni* have been observed to inhabit the abandoned burrows of Persian jirds, *Meriones persicus* (Roberts 1997; Hamidi et al. 2017). The actual suitable habitat of *C. hotsoni* may be broader than predicted by ENM because the model does not adequately include the effects of ecological facilitation. Hamidi et al. (2019) produced an ENM model that differs slightly from this study even though

the same localities were used. They find suitability in eastern Iran is high and suitability is moderately good throughout much of southern Iran and Pakistan. This could be due to different parameters being selected, for example they used 12 bioclimatic parameters compared to 5 in this study. Effects of ecological facilitation cannot be measured with climatic parameters so each model might be missing an important key when predicting distributions.

Overfitting and underfitting are common mistakes when inputting parameters into ENMs where models that are produced can be either too complex or not detailed enough to accurately predict the distribution the user is trying to produce (Radosavljevic and Anderson 2014). Small sample sizes increase the potential for overfitting with ENM, but Anderson and Gonzalez (2011) demonstrated that if parameters are fit to specific species that models can be greatly improved. Even if the model for *C*. sp. Khy were not fit correctly, the model for *C. baluchi* agrees that there is a distinction in the ecology between the two. The *C. baluchi* ENM model is based on a larger pool of samples and is most likely a better model than the *C*. sp. Khy model that had very low sampling. At this point, the support for *C*. sp. Khy as a distinct species includes *Cytb*, *Rbp3*, genomic analysis, and distinct ecological requirements.

ACKNOWLEDGEMENTS

This project was developed with the support and guidance of Dr. Ryan Norris, without whom I would not have had the opportunity to work with these samples. I would like to thank Ola Balkowiec for assistance with extractions. Pakistan samples were generously provided by C. William Kilpatrick and Charles A. Woods. Also, I would like to thank Jamshid Darvish and Jim Patton for providing the Iranian samples. Thank you to other collaborators on this project: Brooke Rawson, Andreas Chavez, Kordiyeh Hamidi, and Safie Akbarirad. In addition, the feedback and encouragement by the Norris Lab, especially Rachel Crites, was greatly appreciated.

LITERATURE CITED

Akbarirad, S., Darvish, J., Aliabadian, M. 2016. Increased species diversity of brush-tailed mice, genus *Calomyscus* (Calomyscidae, Rodentia), in the Zagros Mountains, western Iran. Mammalia 80:549-562.

Akbarirad, S., Darvish, J., Aliabadian, M. 2018. Taxonomic research on *Calomyscus baluchi* from Bamiyan in Afghanistan and molecular comparison with *C. baluchi* from Pakistan. Journal of Research in Biology1(1),e1518.

Anderson, R.P., and Gonzalez Jr., I. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. Ecological Modeling 222:2796-2811.

Beck, H., Zimmermann, N., McVicar, T., Vergopolan, N., Berg, A., and Wood, E.F. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data 5: 180214.

Eskandarzadeh, N., Rastegar-Pouyani, N., Rastegar-Pouyani, E., Fathinia, B., Bahmani, Z., Hamidi, K., and Gholamifard, A. 2018. Annotated checklist of the endemic Tetrapoda species of Iran. Zoosystema 40. 10.5252/zoosystema2018v40a24.

Feng, X., Park, D.S., Walker, C., Peterson, A.T., Merow, C., and Papes, M. 2019. A checklist for maximizing reproducibility of ecological niche models. Nature Ecology and Evolution 3:1382-1395.

Hamidi, K., Darvish, J., Matin, M.M. 2017. Ecological survey of two Calomyscidae species; Goodwin's brush-tailed mouse and Hotson's brush-tailed mouse (Rodentia) in the eastern parts of Iran. Acta Ecologica Sinica 37:2, 105-114.

Hamidi, K., Matin, M.M., Kilpatrick, C.W., and Eskandarzadeh, N. 2019. Landscape and niche specialisation of two brush-tailed mice species *Calomyscus elburzensis* and *C. hotsoni* in Iran: a case of the role of ecological niche modelling in finding area(s) of contact, Ethology Ecology & Evolution 31:5, 435-456.

Jansa S.A., Voss R.S. 2000. Phylogenetic studies on didelphid marsupials I. Introduction and preliminary results from nuclear IRBP gene sequences. Journal of Mammalian Evolution 7:43–77.

Kamp, U. and Owen, L.A. 2011. Late quaternary glaciation of Northern Pakistan. Developments in Quaternary Science 15:909-927.

Kass, J. M., Meenan, S. I., Pinilla-Buitrago, G. E., Merow, C., Anderson, RP. 2018b. Wallace ecological modeling application vignette. https://wallaceecomod.github.io/vignettes/wallace_vignette.html Accessed 7 November 2019.

Kass J. M., Vilela B., Aiello-Lammens M. E., Muscarella R., Merow C., Anderson R.P., 2018a. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. Methods in Ecology and Evolution 9:1151–1156.

Kilpatrick, C.W. 2002. Noncryogenic preservation of mammalian tissue for DNA extraction: An assessment of storage methods. Biochemical Genetics 40:53-62.

Kilpatrick, C.W. 2017. Family Calomyscidae (Brush-tailed Mice) Pp. 144-155 in D.E. Wilson, T.E Lacher, and R.A. Mittermeier. Handbook of Mammals of the World Vol. 7. Lynx Edicions.

Merlone, A., Al-Dashti, H., Faisal, N., Cerveny, R.S., Alarmi, S., Bessemoulin, P., Brunet, M., Driouech, F., Khalatyan, Y., Peterson, T.C., Rahimzadeh, F., Trewin, B., Wahab, M.M.A., Yagan, S., Coppa, G., Smorgon, D., Musacchio, C., Krahenbuhl, D. 2019. Temperature extreme records: World Meteorological Organization metrological and meteorological evaluation of the 54.0°C observations in Mitribah, Kuwait and Turbat, Pakistan in 2016/2017. International Journal of Climatology 39:5154-5169.

Merow C., Smith M.J., Silander J.A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. Ecography 36: 1058-1069.

Musser, G. G., and Carleton M. D. 2005. Mammal species of the world: a taxonomic and geographic reference. 3rd edition (D. E. Wilson and D. M. Reeder, eds.). Johns Hopkins University Press. Baltimore, Maryland.

Nix, H.A. 1986. BIOCLIM, a bioclimatic analysis and prediction system. BIOCLIM - A Bioclimatic Analysis and Prediction System 59.

Norris, R.W., Rawson, B., Kilpatrick, C.W., Akbarirad, S., Weller, A.K., Woods, C.A., and Chavez, A.S. In prep. Extensive hidden diversity in the brush-tailed mice, genus *Calomyscus* (Rodentia; Calomyscidae), is indicated by mitochondrial and ddRADseq data. To be submitted to Journal of Mammalogy.

Norris, R.W., Morshed, S., Kilpatrick, C.W., Woods, C.A., Perelman, P.L., Romanenko, S.A., and Malikov, V.G. 2003. The new data on the diversity of *Calomyscus* Thomas, 1905 (Rodentia, Calomyscinae). Pp 166-169 in Proceedings of International Conference devoted to the 90th anniversary of Prof. I.M. Gromov. Saint Petersburg.

Norris, R.W., Woods, C.A., and Kilpatrick, C.W. 2008. Morphological and molecular definition of *Calomyscus hotsoni* (Rodentia: Muroidea: Calomyscidae). Journal of Mammalogy 89:306-315.

O'Donnell, M.S., and Ignizio, D.A., 2012, Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691, 10 p.

Peel, M. C., Finlayson, B. L., and McMahon, T. A. 2007. Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences 11: 1633–1644.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7: e37135.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. and Blair, M. E. 2017. Opening the black box: an open-source release of Maxent. Ecography 40: 887–893.

Rawson, B. 2019. Phylogenomics and species delimitation in the brush-tailed mouse genus *Calomyscus* using ddRADSeq data. Undergraduate Thesis at The Ohio State University.

Roberts, T.J. 1977. The mammals of Pakistan. Oxford University Press, Karachi.

Roberts, T.J. 1997. The mammals of Pakistan, revised edition. Oxford University Press, Oxford.

Schenk, J. J., Rowe, K.C., Steppan, S.J. 2013. Ecological Opportunity and Incumbency in the Diversification of Repeated Continental Colonizations by Muroid Rodents, Systematic Biology 62:6, 837–864.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

Toews, D.P. and Brelsford, A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology 21:3907-3930.

Vorontsov, N. N., Kartavtesva, I. V., and Potapova, E. G. 1979. Systematics of the genus *Calomyscus*. 1. Karyological differentiation of the sibling species from Transcaucasia and Turkmenia and a review of species of the genus *Calomyscus*. (in Russian). Zoologischaskii Zhurnal 58: 1215-1224.

FIGURE LEGENDS

FIGURE 1- Geographic distributions of the *C. bailwardi* species group. Shaded ranges displayed on map represent current distributions based on *Cytb* data. Points on map represent locations of *Rbp3* sampling for this study. The box situated in the top center is a parsimony network constructed of the *Rbp3* results. Size of circle correlates to number of samples in haplotype. Lines between circles denotes one difference in sequence, as well as each hatch mark.

FIGURE 2- Midpoint rooted phylogenetic tree constructed using maximum parsimony in PAUP* with *Rbp3* data. Tissue numbers correspond to University of Vermont Tissues (Appendix A).

FIGURE 3- Model of *C. baluchi*. a) Potential suitability across *C. bailwardi* species group distribution. b) Corresponding response curves.

FIGURE 4- Model of *C*. sp. Khy. a) Potential suitability across *C. bailwardi* species group distribution. b) Corresponding response curves.

FIGURE 5- Model of *C. hotsoni*. a) Potential suitability across *C. bailwardi* species group distribution. b) Corresponding response curves.

Figure 1.

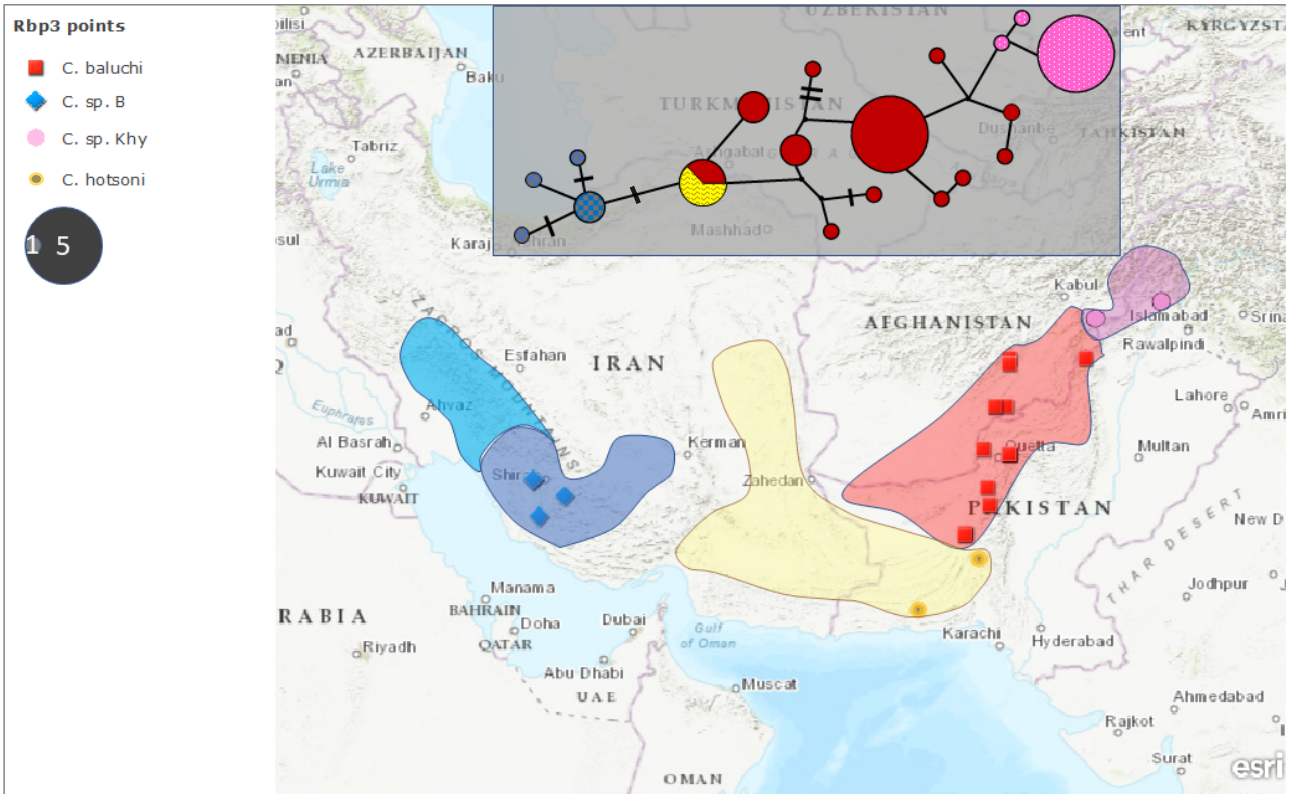

Rbp3 points
C. baluchi
C. sp. B
C. sp. Khy
C. hotsoni
1 5
AZERBAIJAN
Baku
Tabriz
Lake Urmia
Karaj
Mashhad
Kabul
AFGHANISTAN
Islamabad
Rawalpindi
Lahore
Multan
Quetta
Esfahan
IRAN
Ahvaz
Euphrates
Al Basrah
Kuwait City
KUWAIT
Kerman
Zahedan
PAKISTAN
THAR DESERT
Jodhpur
Manama
BAHRAIN
Doha
QATAR
Dubai
Riyadh
Abu Dhabi
UAE
Gulf of Oman
Muscat
Karachi
Hyderabad
Ahmedabad
Rajkot
Surat
OMAN
esri

Figure 2

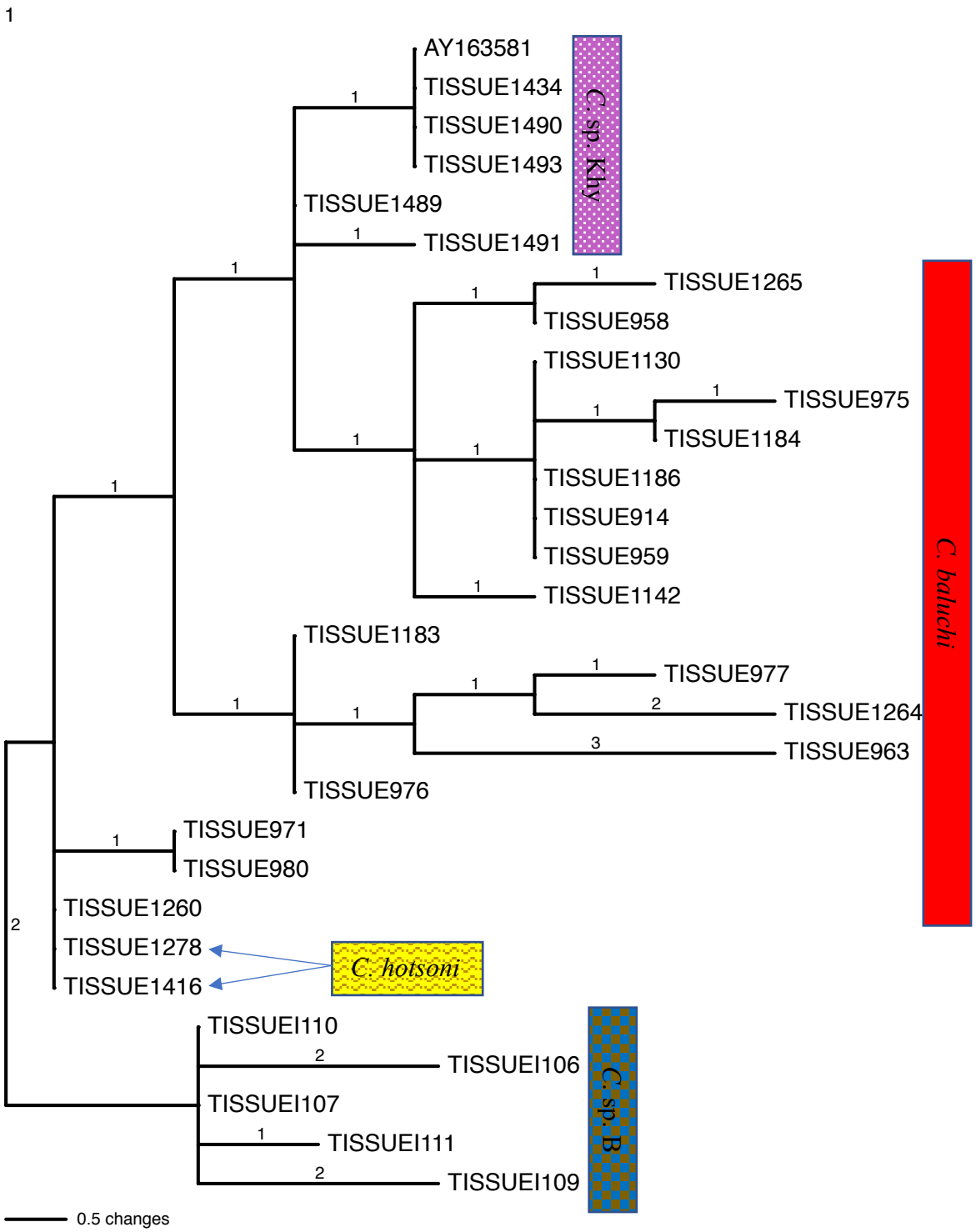
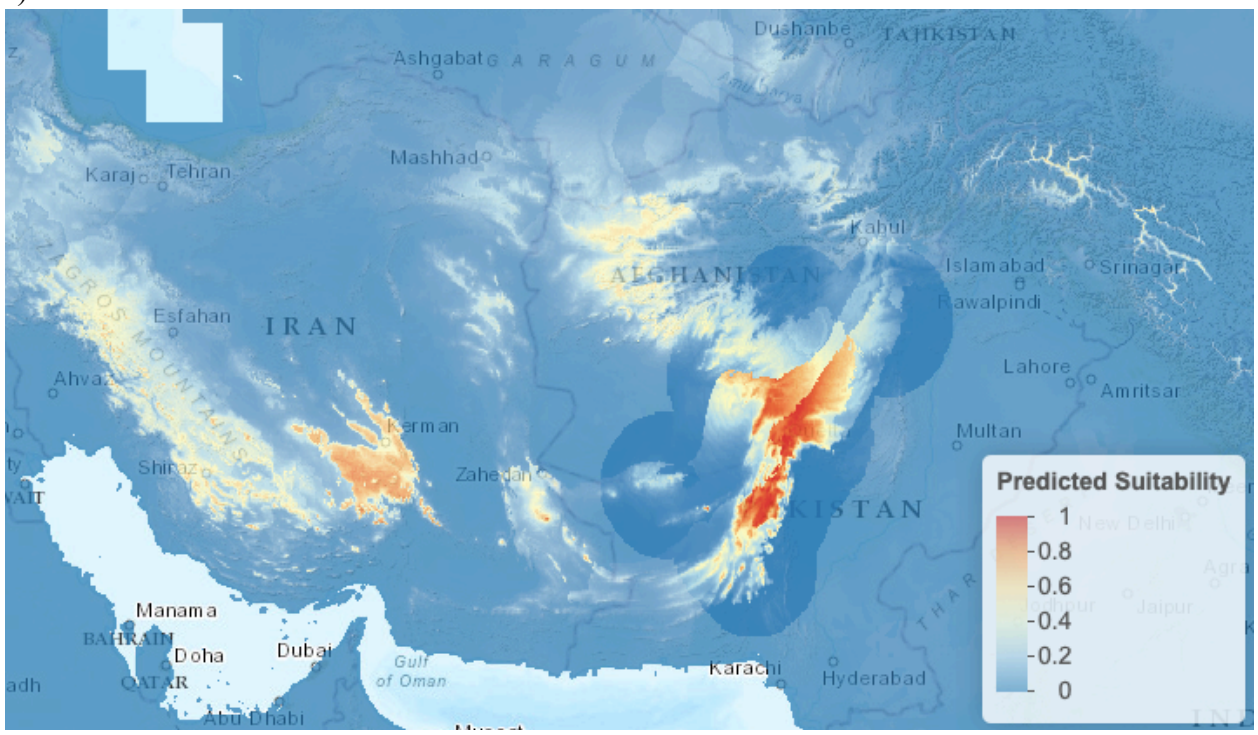

AY163581
TISSUE1434
TISSUE1490
TISSUE1493
TISSUE1489
TISSUE1491
C. sp. Khy
TISSUE1265
TISSUE958
TISSUE1130
TISSUE975
TISSUE1184
TISSUE1186
TISSUE914
TISSUE959
TISSUE1142
C. baluchi
TISSUE1183
TISSUE977
TISSUE1264
TISSUE963
TISSUE976
TISSUE971
TISSUE980
TISSUE1260
TISSUE1278
TISSUE1416
C. hotsoni
TISSUEI110
TISSUEI106
TISSUEI107
TISSUEI111
TISSUEI109
C. sp. B
0.5 changes

Figure 3. *C. baluchi*

a)



b)

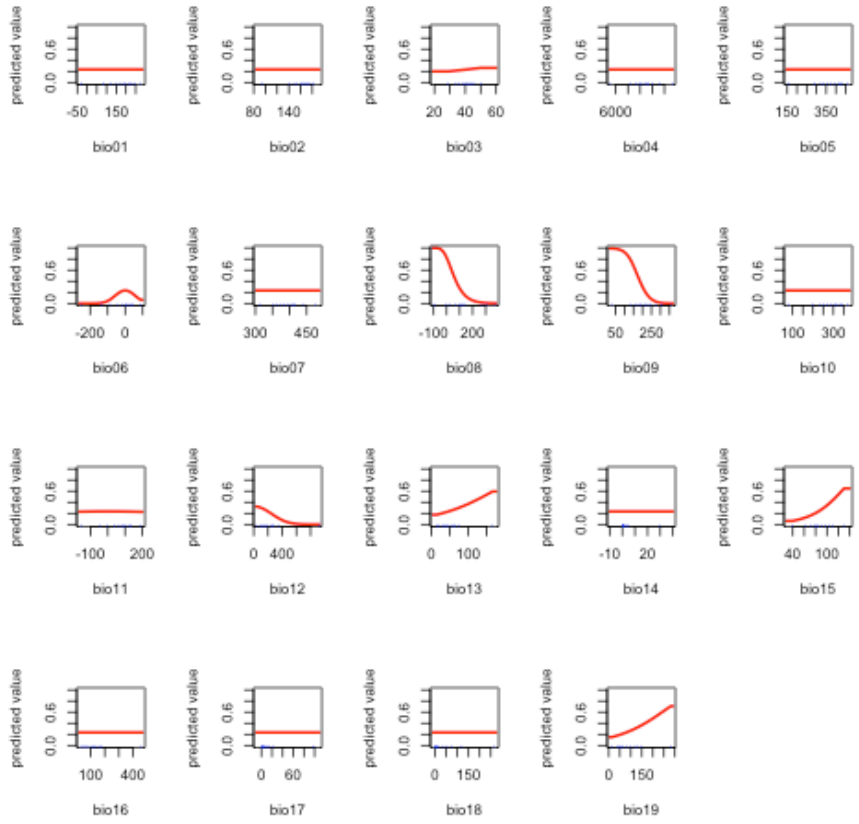


Figure 4. *C.* sp. Khy

a)

b)

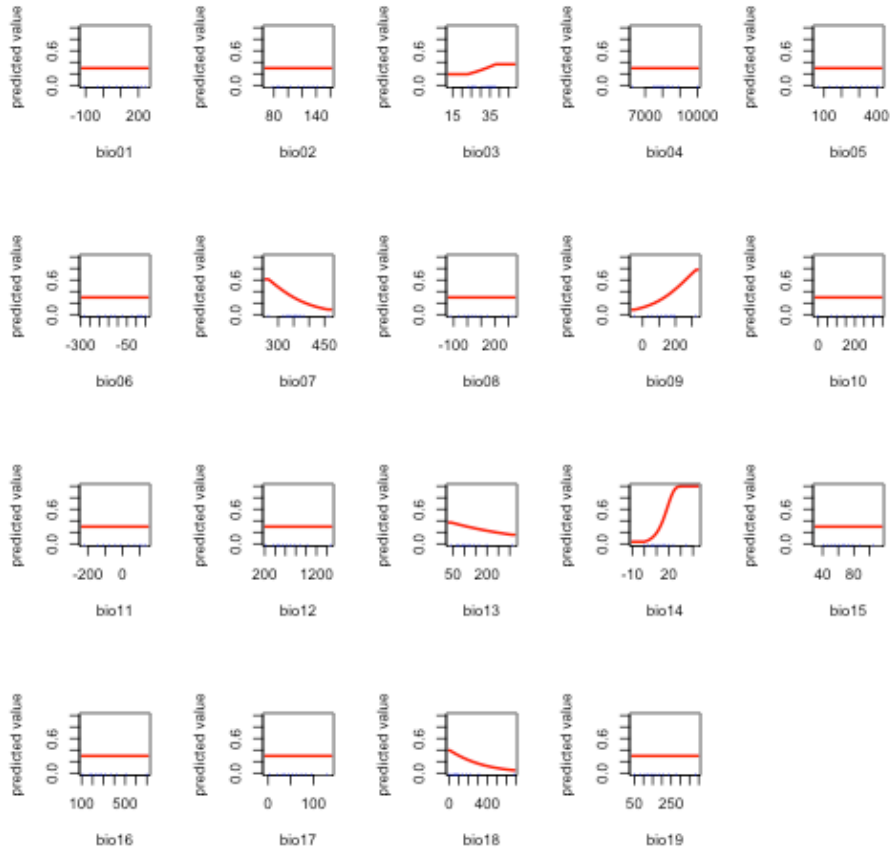
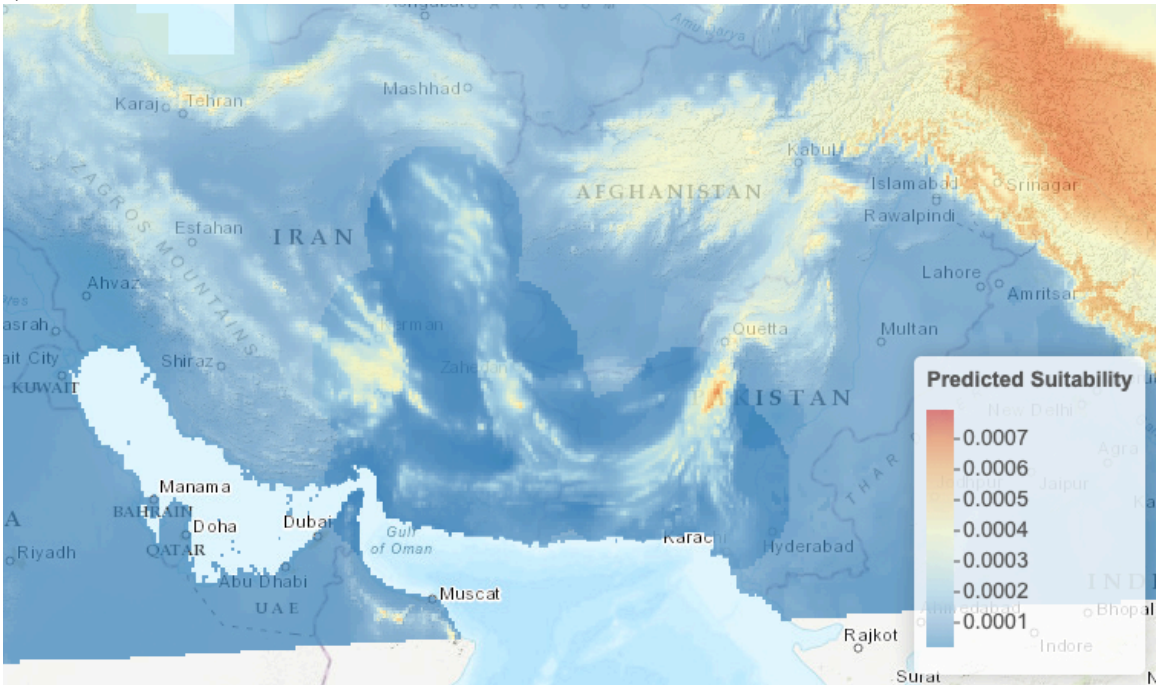
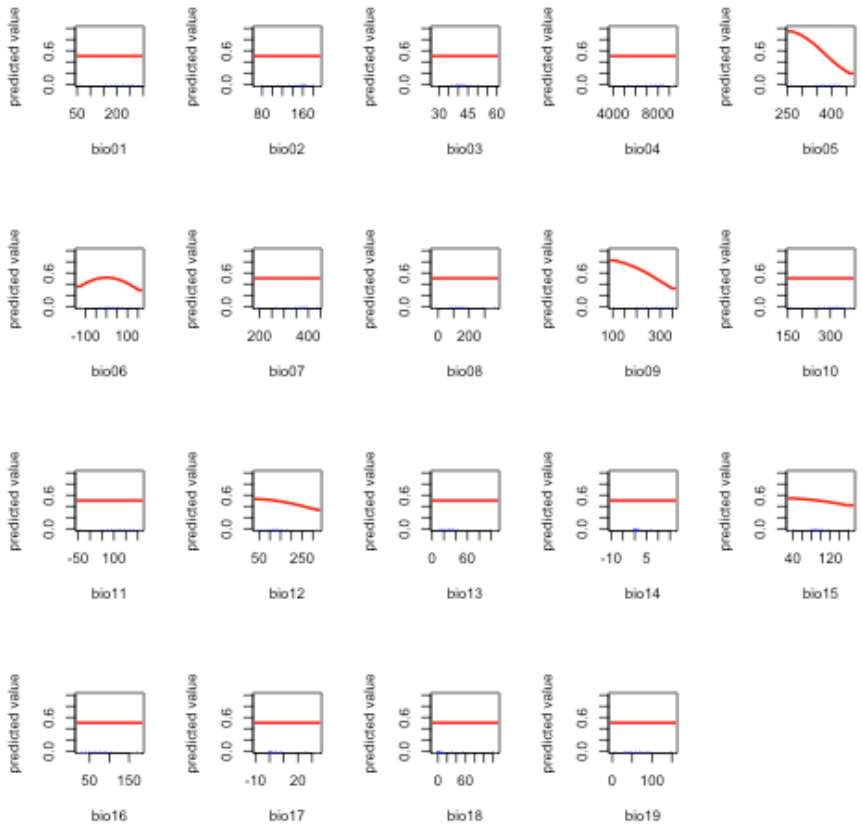


Figure 5. *C. hotsoni*

a)



b)



Appendix A- *Rbp3* Network Tissues

Species	Tissue	Museum	Locality	Latitude (N)	Longitude (E)
C. baluchi	UVM 1130	FLMNH 28998	Pakistan, Balochistan, Harboi Hills, Khan's Bungalo	28.88	66.71
C. baluchi	UVM 1142	FLMNH 28864	Pakistan, Balochistan, 38.9 km NE Kalat, 2600 m (type locality)	29.36	66.64
C. baluchi	UVM 1183	FLMNH 28881	Pakistan, Balochistan, Zhob, Shinghar, 2600 m	31.6	67.2333333
C. baluchi	UVM 1184	FLMNH 28882	Pakistan, Balochistan, Zhob, Shinghar, 2600 m	31.6	66.8666667
C. baluchi	UVM 1186	FLMNH 28884	Pakistan, Balochistan, Zhob, Shinghar, 2600 m	31.6	66.8666667
C. baluchi	UVM 1260	FLMNH 28993	Pakistan, Khyber Pakhtunkhwa, N. Waziristan, Datta Khel	32.8833333	67.3305556
C. baluchi	UVM 1264	FLMNH 28997	Pakistan, Khyber Pakhtunkhwa, N. Waziristan, Datta Khel	32.8833333	69.7666667
C. baluchi	UVM 1265	FLMNH 28998	Pakistan, Khyber Pakhtunkhwa, S. Waziristan, Razmak	32.75	67.3305556
C. baluchi	UVM 914	FLMNH 28472	Pakistan, Balochistan, Ziarat, Fern Tangi, 2350 m	30.42	66.5
C. baluchi	UVM 958	FLMNH 28513	Pakistan, Balochistan, Kalat, Kalgali, 1790 m	28.0780556	65.9319444
C. baluchi	UVM	FLNHM	Pakistan, Balochistan, Kalat,	28.0780556	65.9319444

	959	28514	Kalgali, 1790 m		
C. baluchi	UVM 963	FLMNH 28516	Pakistan, Balochistan, Kharan, Kalgali, 1560 m	28.0272222	65.9116667
C. baluchi	UVM 971	FLMNH 28488	Pakistan, Balochistan, Sibi, Toreshore, 2360 m	30.2961111	67.3305556
C. baluchi	UVM 975	FLMNH 28491	Pakistan, Balochistan, Sibi, Toreshore, 2360 m	30.2961111	67.3305556
C. baluchi	UVM 976	FLMNH 28492	Pakistan, Balochistan, Sibi, Toreshore, 2360 m	30.2961111	67.3305556
C. baluchi	UVM 977	FLMNH 28493	Pakistan, Balochistan, Sibi, Toreshore, 2360 m	30.2961111	67.3305556
C. baluchi	UVM 980	FLMNH 28485	Pakistan, Balochistan, Sibi, Toreshore, 2360 m	30.2961111	67.3305556
C. hotsoni	UVM 1278	FLMNH 29022	Pakistan, Balochistan, Panjgur, Mitho Singh, 1440 m	25.8969444	64.4302778
C. hotsoni	UVM 1416	FLMNH 29786	Pakistan, Balochistan, Khzoler, Wad	27.35	66.37
C. sp. B	Tissue I106		2 km W of Khollar, prov. Fars, alt. 2,130 m	29.59	52.12
C. sp. B	Tissue I107		2 km W of Khollar, prov. Fars, alt. 2,130 m	29.59	52.12
C. sp. B	Tissue I109		5 km SE of Posht Chenar, prov. Fars, alt. 1,780 m	29.13	53.12

C. sp. B	Tissue I110		2 km W of Khollar, prov. Fars, alt. 2,130 m	29.59	52.12
C. sp. B	Tissue I111		10 km N of Firuzabad, prov. Fars, alt. 1,410 m	28.56	52.31
C. sp. Khy	AY163581	FMNH140412	Pakistan, Malakand Div, Swat Dist, Oriental, Karakar Pass		
C. sp. Khy	UVM 1434	FLMNH 29702	Pakistan, Khyber Pakhtunkhwa, Swat, Karakas Pass, 1311 m	34.4	72.2
C. sp. Khy	UVM 1489	FLMNH 29685	Pakistan, Khyber Pakhtunkhwa, Kurram, Makay Village, 2250 m	33.95	70.1
C. sp. Khy	UVM 1490	FLMNH 29686	Pakistan, Khyber Pakhtunkhwa, Kurram, Makay Village, 2250 m	33.95	70.1
C. sp. Khy	UVM 1491	FLMNH 29687	Pakistan, Khyber Pakhtunkhwa, Kurram, Makay Village, 2250 m	33.95	70.1
C. sp. Khy	UVM 1493	FLMNH 29692	Pakistan, Khyber Pakhtunkhwa, Kurram, Makay Village, 2250 m	33.95	70.1

Appendix B-Ecological Niche Modeling localities

Species	Longitude	Latitude
C. baluchi	64.1	29.33
C. baluchi	64.12	29.31
C. baluchi	65.78	31.66
C. baluchi	65.91	28.03
C. baluchi	65.93	28.08
C. baluchi	66.38	27.55
C. baluchi	66.5	29.17
C. baluchi	66.5	30.42
C. baluchi	66.5	31.6
C. baluchi	66.64	29.36
C. baluchi	66.71	28.88
C. baluchi	66.87	30.03
C. baluchi	66.87	30.42
C. baluchi	66.87	29.52
C. baluchi	66.87	31.6
C. baluchi	67.23	29.88
C. baluchi	67.23	30.42
C. baluchi	67.23	31.6
C. baluchi	67.33	30.3
C. baluchi	67.33	29.93
C. baluchi	67.33	30.52
C. baluchi	67.33	31.27
C. baluchi	67.33	31.2
C. baluchi	67.33	31.6
C. baluchi	67.33	32.75
C. baluchi	67.33	32.88
C. baluchi	67.67	30.42
C. baluchi	67.67	31.2
C. baluchi	67.73	30.37
C. baluchi	67.73	31.2
C. baluchi	69.77	32.88
C. hotsoni	57.19	29.33

C. hotsoni	57.24	30.04
C. hotsoni	57.49	26.5
C. hotsoni	58.04	28.6
C. hotsoni	58.69	32.92
C. hotsoni	58.86	32.44
C. hotsoni	58.91	32.41
C. hotsoni	59.21	32.81
C. hotsoni	59.64	26.55
C. hotsoni	60.36	26.39
C. hotsoni	60.78	29.74
C. hotsoni	60.85	29.51
C. hotsoni	61	28.34
C. hotsoni	61.17	28.15
C. hotsoni	61.25	28.35
C. hotsoni	61.46	27.18
C. hotsoni	61.77	27.3
C. hotsoni	64.09	26.96
C. hotsoni	64.15	26.76
C. hotsoni	65.84	28.02
C. hotsoni	66.83	26.73
C. hotsoni	66.88	26.38
C. sp. Khy	70.1	33.95
C. sp. Khy	70.1	33.96
C. sp. Khy	71.43	35.34
C. sp. Khy	71.77	35.83
C. sp. Khy	71.77	35.73
C. sp. Khy	72.29	34.6

Appendix C- WorldClim Global Climate Data

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (BIO5-BIO6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter